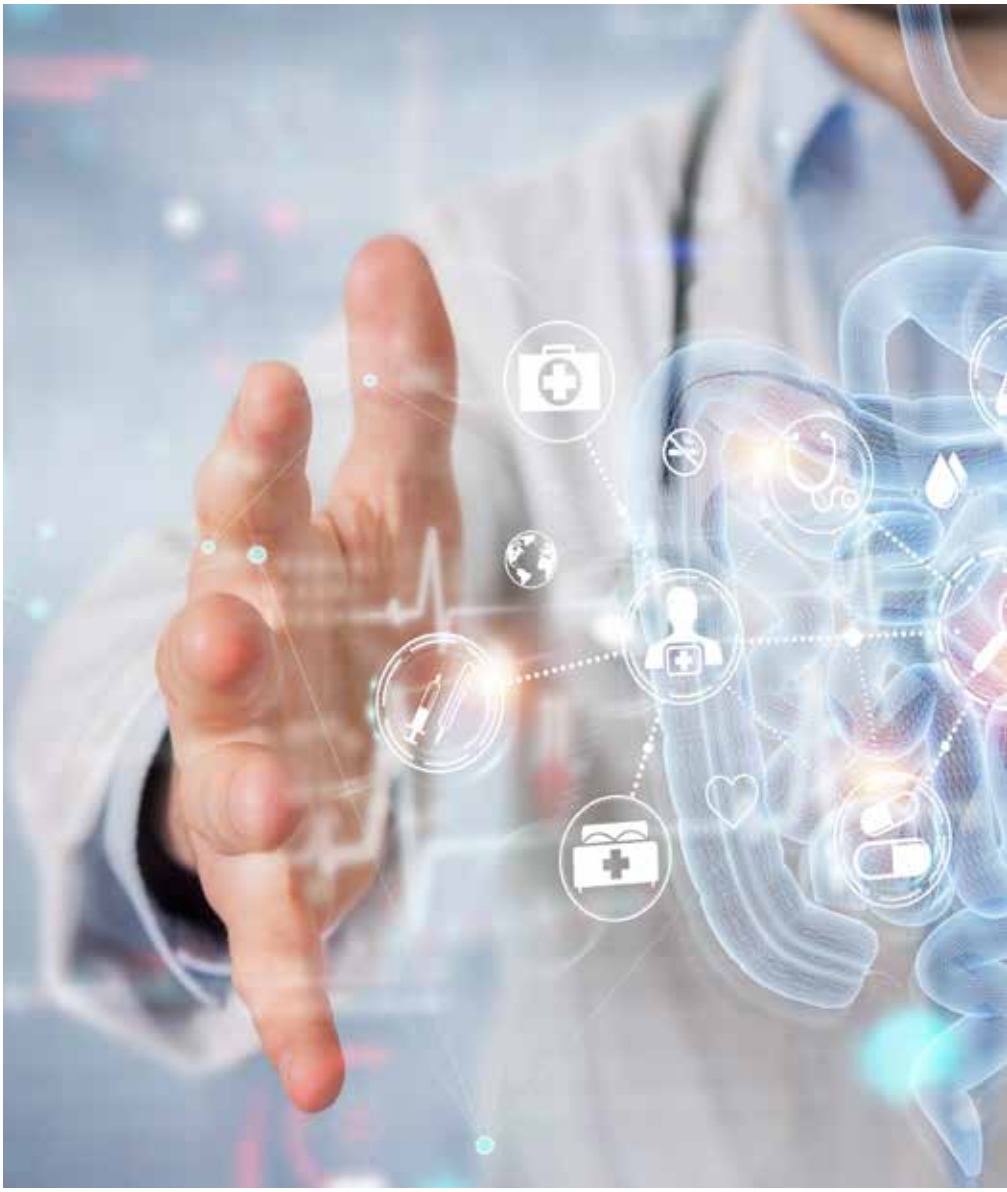
Artificial Intelligence in Gastroenterology

AI is becoming increasingly ingrained in the biopharma sector, and it is no different within gastroenterology – but what might be the challenges involved in its usage?

Dr Marcela Vieira at Clario

AI is advancing quickly in the gastrointestinal (GI) field. Recent discussions are focused on how these tools can be implemented into clinical practice, clinical training, and, moreover, into the clinical trials space; transforming recruitment, drug discovery, patient safety, and data collection, including image acquisition and interpretation.



Central reading is a decisive component in the analysis of trials that involve medical imaging, and consists of an independent and blinded review of colonoscopy video images by trained physicians.

In contrast to other areas of medicine such as oncology and cardiology – where the imaging endpoints have historically relied on central reading – gastroenterology and, more specifically, Inflammatory Bowel Disease (IBD) trials traditionally had interpretation of endoscopic findings only by local physicians. The issue with these assessments is the non-blinded status of the observer (1).

One study compared results from a mesalamine trial in Ulcerative Colitis (UC) where the initial analysis showed no statistical difference between treatment and placebo. The endoscopic videos were later reassessed by a blinded central reader. The post hoc analysis showed that the study would have demonstrated statistically significant efficacy had some subjects been excluded for not meeting

VISIT…

eligibility criteria, over the central reader evaluation (2).

There are several scoring systems available for tracking the disease state in patients with UC, but currently the Mayo score is the most widely used because it incorporates endoscopic and patient-reported symptoms (3).

Scoring colonoscopy videos of UC patients requires a high level of expertise, but even among trained expert central readers there are often disagreements. Although the output of the endoscopic Mayo score is numeric, it relies on subjective interpretation from physicians of the colonoscopy findings which is also highly dependent on the technical procedures followed, such as quality of bowel preparation, proper washing, distance from mucosa, insufflation, and withdrawal time. Human readers may also face fatigue and burnout that can result from spending too much time in front of a screen reviewing colonoscopy videos. In clinical trials, a single score can determine a patient's eligibility into a study, meaning an inaccurate score could negatively impact both subject selection and assessment of treatment response.

There are multiple mitigation strategies used in central reading to avoid errors such as reader training, monitoring, and reading paradigms based on multiple readers and an adjudicator (4). These may help decrease variability, and increase accuracy and reliability of the reads.

It has been shown that consistency among central readers might also be improved by AI algorithms that automatically process videos to identify salient features and estimate the level of disease activity using established scoring methods (5).

As AI technology in the gastrointestinal field continues to advance, there is a race to apply these tools in clinical trials to improve data reading efficiency and realise cost savings. The biggest challenge is access to large, expertly annotated endoscopic video libraries that are required to train and test the algorithms. One obvious source of videos is the numerous completed clinical trials provided that subject consent allows this expanded use of the collected study videos. Gotlieb *et al* suggest a consortium formed by pharmaceutical companies, working in a pre-competitive space, to supply those videos to be reread by experts (6). Some key aspects of the algorithm are the variability of the dataset, any inherent bias from the reviewers, and if it has been tested in clinical trials.

Qualification of the reviewers and inherent bias are concerns while determining the ground truth. Some readers may be clinical and disease experts but are not adequately trained on scoring systems, or may have trained in different parts of the world where scoring patterns might be divergent. Heterogeneity of the dataset is also an integral part of the AI algorithm, as it must be reproducible and applicable for larger populations and videos of varying quality.

There is no question that AI should be embraced as a complement to central reading. It is an open question how best to leverage the speed and inherent objectivity of computer algorithms and improve the quality of the central reading, while not losing oversight of the human reader.

Currently the central read paradigm more commonly used in UC clinical trials is a model where two independent

readers assess a case, and an adjudicator reviews the case if there is disagreement between the primary readers ('2+1' read paradigm). This process improves the accuracy of the results, but generally increases the costs and turnaround times for results (7, 8).

Predicting a central reader score on video recordings from subjects with Ulcerative Colitis with the help of AI could decrease both time and cost, while sustaining similar (or even better) quality and accuracy. Although it has not yet been applied in a clinical trial, some experts are speculating that the addition of a AI-predicted score could be positively integrated into a 2+1 paradigm, where the human reader would still act as an adjudicator.

AI tools in GI also bring regulatory challenges, including data sources for the algorithms and how they are used, risks of AI failures, and the ethical considerations if these have an impact on patient health outcomes. To address these challenges, the International Coalition of Medicines Regulatory Authorities (ICMRA) set up some recommendations, such as a risk-based approach to assessing and regulating AI, governance structures in industry to oversee algorithms, and regulatory guidelines for development and validation (9).

In clinical practice there are other key aspects, one being how current training programmes can educate younger physicians and fellows in AI and computer science, and how to best integrate that into the clinical practice. This is an era where computer science and understanding technology has become critically important.

The new generation of gastroenterologists need to be prepared to understand AI terminology and basics of ML algorithms, and in an academic setting, to collaborate with computer science engineers. A formal training on the readily available AI tools (e.g., polyp detection technologies) will become necessary so physicians are able to implement those into their day-to-day activities.

Predictive algorithms, such as the ones being developed for the Mayo score, will also be a useful tool in the GI training programmes, helping trainees to identify lesions and how to score them properly. Further comparisons with the AI scores could also help understand fellows' preparedness for making diagnosis and evaluation of performance for senior faculty.

A recurring concern is how the use of auxiliary methods such as lesion detection tools could interfere with the physician's 'trained eyes' to evaluate colonoscopies, when these methods are not available. While descaling might be a valid concern, it is also part of the evolving process of AI not only in GI, but in medicine in general. There is no evidence that AI may decrease a clinician's capability of detecting lesions, while also having the added benefit of providing real-time 'feedback' for the observer, who may learn from the process. AI also brings along the capability of reducing error so there is less risk of missing an important feature.

There is a limited number of physicians skilled in computer science and vice-versa. Therefore, there is an urgent need for cross-collaboration in this field. These physician-scientist partnerships have a great potential in their hands if they work synergistically. Indeed, these collaborations are already bringing tremendous changes to the GI endoscopy field and ultimately benefiting the patients, which should always be kept in the forefront of these advances.

References

1. *Panes J et al, Central Endoscopy Reading in Inflammatory Bowel Diseases, Journal of Crohn's and Colitis, pp S542–S547, 2016*
2. *Feagan BG et al, The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis, Gastroenterology; 145: pp149-57, 2013.*
3. *Mohammed Vashist et al, Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis, Cochrane Database Syst Rev 1:CD011450, 2018.*
4. *Ahmad H et al, Central endoscopy reads in inflammatory bowel disease clinical trials: The role of the imaging core lab. Gastroenterology Report; pp 1-6, 2014*
5. *Kani HT et al, Evaluation of endoscopic mayo score with an artificial intelligence algorithm, Journal of Crohn's and Colitis, pp S195-S196, 2021.*
6. *Gottlieb K et al, Endoscopy and central reading in inflammatory bowel disease clinical trials: achievements, challenges, and future developments Gut, 70: pp 418-426, 2021.*
7. *Ahmad HA et al, The 2+1 paradigm: an efficient algorithm for central reading of Mayo endoscopy subscores in global multicenter phase 3 ulcerative colitis clinical trials. Gastroenterology Report 4: pp35-38, 2015*
8. *Gottlieb K et al, Voting for image scoring and assessment (VISA) – theory and application of a 2+1 reader algorithm to improve accuracy of imaging endpoints in clinical trials, BMD Medical Imaging 15: pp 6-10, 2015*
9. *Visit: www.ema.europa.eu/en/news/artificial-intelligence-medicine-regulation*

Dr Marcela Vieira is a physician-scientist and gastrointestinal endoscopist responsible for the Gastroenterology Division at **Clario** where, as Medical Director, she leads the design and medical and scientific oversight of the imaging endpoints for clinical trials.

Marcela worked at AstraZeneca as a Clinical Research Physician and as a Gastroenterology Researcher at Johns Hopkins University, US. She has experience from early and late phase trials, having worked on Drug Safety and Clinical Development teams across various programmes in immunology.

Now, her focus is on innovation and collaboration, bringing cutting edge imaging technology to the care of patients with gastrointestinal disorders.